



Contrasting Genetic Diversity of *Listeria* Pathogenicity Islands 3 and 4 Harbored by Nonpathogenic *Listeria* spp.

 Sangmi Lee,^a  Cameron Parsons,^b Yi Chen,^c Robert S. Dungan,^d Sophia Kathariou^b

^aDepartment of Food and Nutrition, Chungbuk National University, Cheongju, Chungbuk, South Korea

^bDepartment of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Raleigh, North Carolina, USA

^cDivision of Microbiology, Center for Food Safety and Applied Nutrition, Food and Drug Administration, College Park, Maryland, USA

^dNorthwest Irrigation and Soils Research Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Kimberly, Idaho, USA

ABSTRACT *Listeria monocytogenes* causes the severe foodborne disease listeriosis. Several clonal groups of *L. monocytogenes* possess the pathogenicity islands *Listeria* pathogenicity island 3 (LIPI-3) and LIPI-4. Here, we investigated the prevalence and genetic diversity of LIPI-3 and LIPI-4 among 63 strains of seven nonpathogenic *Listeria* spp. from the natural environment, i.e., wildlife (black bears [*Ursus americanus*]) and surface water. Analysis of the whole-genome sequence data suggested that both islands were horizontally acquired but differed considerably in their incidence and genetic diversity. LIPI-3 was identified among half of the *L. innocua* strains in the same genomic location as in *L. monocytogenes* (*guaA* hot spot) in a truncated form, with only three strains harboring full-length LIPI-3, and a highly divergent partial LIPI-3 was observed in three *Listeria seeligeri* strains, outside the *guaA* hot spot. Premature stop codons (PMSCs) and frameshifts were frequently noted in the LIPI-3 gene encoding listeriolysin S. On the other hand, full-length LIPI-4 without any PMSCs was found in all *Listeria innocua* strains, in the same genomic location as *L. monocytogenes* and with ~85% similarity to the *L. monocytogenes* counterpart. Our study provides intriguing examples of genetic changes that pathogenicity islands may undergo in nonpathogenic bacterial species, potentially in response to environmental pressures that promote either maintenance or degeneration of the islands. Investigations of the roles that LIPI-3 and LIPI-4 play in nonpathogenic *Listeria* spp. are warranted to further understand the differential evolution of genetic elements in pathogenic versus nonpathogenic hosts of the same genus.

IMPORTANCE *Listeria monocytogenes* is a serious foodborne pathogen that can harbor the pathogenicity islands *Listeria* pathogenicity island 3 (LIPI-3) and LIPI-4. Intriguingly, these have also been reported in nonpathogenic *L. innocua* from food and farm environments, though limited information is available for strains from the natural environment. Here, we analyzed whole-genome sequence data of nonpathogenic *Listeria* spp. from wildlife and surface water to further elucidate the genetic diversity and evolution of LIPI-3 and LIPI-4 in *Listeria*. While the full-length islands were found only in *L. innocua*, LIPI-3 was uncommon and exhibited frequent truncation and genetic diversification, while LIPI-4 was remarkable in being ubiquitous, albeit diversified from *L. monocytogenes*. These contrasting features demonstrate that pathogenicity islands in nonpathogenic hosts can evolve along different trajectories, leading to either degeneration or maintenance, and highlight the need to examine their physiological roles in nonpathogenic hosts.

KEYWORDS *Listeria*, genetic diversity, pathogenicity islands

L *isteria monocytogenes* is a Gram-positive, facultative intracellular bacterial foodborne pathogen responsible for the severe disease listeriosis in humans and other animals, and it is the only human pathogen in the genus *Listeria* (1–3). It is transmitted

Editor Edward G. Dudley, The Pennsylvania State University

Copyright © 2023 American Society for Microbiology. All Rights Reserved.

Address correspondence to Sangmi Lee, sangmilee@chungbuk.ac.kr.

The authors declare no conflict of interest.

Received 14 December 2022

Accepted 6 January 2023

Published 2 February 2023

primarily through ready-to-eat foods that become contaminated in the food processing environment, and populations at especially high risk include the elderly, pregnant women and their fetuses, and immunocompromised individuals (1, 4). Symptoms of systemic listeriosis can be severe, e.g., septicemia, meningitis, stillbirths, and abortions, and case fatality rate is high (16 to 30%), even after administration of antimicrobial treatment (2, 4–7). *L. monocytogenes* is genomically partitioned into four genomic lineages, of which lineages I and II include most strains implicated in human listeriosis (3, 4, 8).

Key to the capacity of *L. monocytogenes* to cause invasive illness is the chromosomal *Listeria* pathogenicity island 1 (LIPI-1), which harbors genes mediating intracellular growth and cell-to-cell dispersal such as *hly*, which encodes the hemolysin listeriolysin O, two different phospholipase genes, *actA*, required for actin polymerization and cell-to-cell spread, and *prfA*, encoding a transcription factor critical for regulation of virulence determinants (9, 10). A similar chromosomal island designated LIPI-2 is harbored by *Listeria ivanovii*, which can cause listeriosis in ruminants (9).

LIPI-1 is harbored by all genomic lineages and serotypes of *L. monocytogenes*, but subsequent investigations revealed two additional *Listeria* pathogenicity islands, designated LIPI-3 and LIPI-4, harbored only by certain *L. monocytogenes* lineages and clones (9–11). Specifically, LIPI-3 is primarily detected in certain strains of lineage I (serotypes 1/2b and 4b) and mediates production of the cytolytic peptide listeriolysin S, expressed exclusively in the intestine and found to have major roles in controlling microbial community composition in the gut lumen (11–14). On the other hand, LIPI-4 was not discovered until multilocus sequence typing (MLST)-based analysis of *L. monocytogenes* clones from human listeriosis and foods revealed that strains of sequence type 4 (ST4) were much more likely to be implicated in human disease than predicted based on their incidence in foods, thus constituting a hypervirulent clone of *L. monocytogenes* (15). Whole-genome sequencing (WGS) analysis of ST4 strains revealed that a cluster of six genes annotated as a cellobiose-family phosphotransferase system (PTS) were harbored by this clone but not by other major clones of *L. monocytogenes* (15). Experimental virulence assessments in the murine model revealed that the genes were critical for neurovirulence and placental virulence of *L. monocytogenes*, and the gene cluster was designated LIPI-4 (15). Even though LIPI-4 was originally considered unique to the hypervirulent clone ST4, subsequent WGS analyses identified LIPI-4 in several other clones, including the emerging clone ST382, implicated in recent outbreaks of listeriosis via produce in the United States (10, 14, 16).

Even though LIPI-3 and LIPI-4 have been extensively investigated in *L. monocytogenes*, our understanding of the evolution of these islands and their distribution in other *Listeria* spp. remains limited. LIPI-3 has been detected in its entirety in several strains of the nonpathogenic species *Listeria innocua*, and these strains could be rendered hemolytic when the LIPI-3-borne *hls* genes were placed under the control of a constitutive promoter (17). On the other hand, analysis of the available literature reveals that LIPI-4 was also harbored by several *L. innocua* strains (10, 18). These studies investigated specific panels of *L. innocua*, primarily from foods, food processing environments, food animals, and silage (17, 18). There is a noticeable dearth of information on incidence or diversity of LIPI-3 or LIPI-4 in *L. innocua* and other nonpathogenic *Listeria* spp. from diverse sources, as evidenced by the fact that these previous studies included just two strains from the natural environment (water puddle and leaves) and two from wildlife, specifically, birds in Finland (17, 18). In this work, we employed WGS-based analysis to investigate the distribution and genomic organization of these pathogenicity islands in a panel of nonpathogenic *Listeria* spp. that included *L. innocua*, *L. seeligeri*, and other *Listeria* spp. that we isolated relatively recently (2014 to 2019) from the aquatic environment and mammalian wildlife (American black bears [*Ursus americanus*]) in the United States.

RESULTS AND DISCUSSION

Genes associated with LIPI-3 are harbored frequently by *L. innocua* of aquatic or wildlife origin but are largely absent from other nonpathogenic *Listeria* spp. Our panel of 63 strains of nonpathogenic *Listeria* spp. included members of seven species,

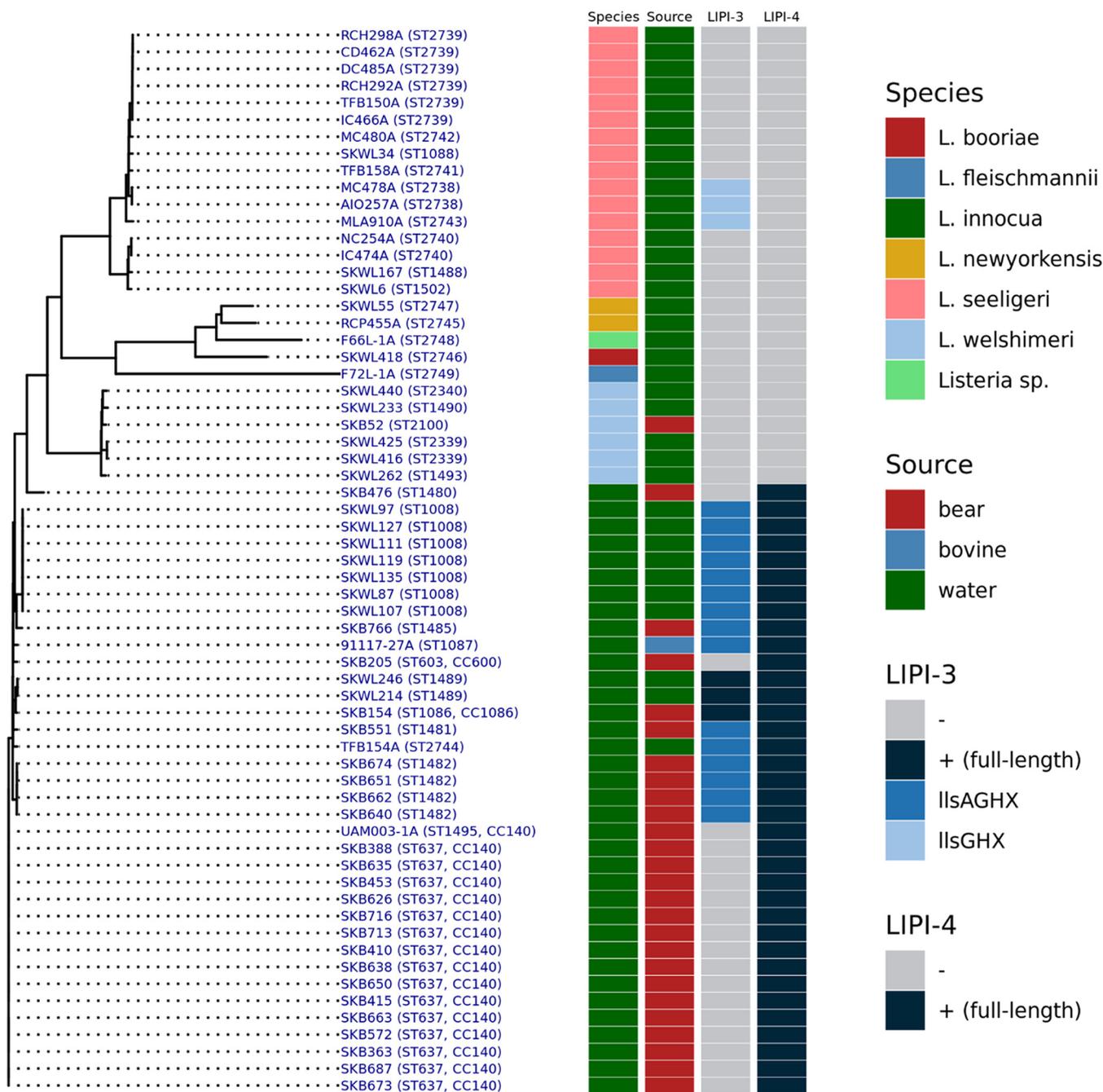


FIG 1 Phylogenetic tree of the whole-genome sequences of the strains investigated in this study. Genome sequences were aligned, bootstrapped, and visualized as described in Materials and Methods. Species, sources, and types of LIPI-3 (absence of LIPI-3 [–], full-length LIPI-3 [+], partial LIPI-3 consisting of *llsAGHX*, and partial LIPI-3 consisting of *llsGHX*) and LIPI-4 (absence of LIPI-4 [–] and full-length LIPI-4 [+]) are shown in the heat map to the right of the tree. ST and, if available, CC designations based on the seven-housekeeping-gene MLST scheme from the Institut Pasteur MLST database are included after the strain designation within parentheses.

of which *L. innocua* ($n = 36$), *L. seeligeri* ($n = 16$), and *L. welshimeri* ($n = 6$) were most predominant (Fig. 1; also, see Table S1 in the supplemental material). Other species included *L. newyorkensis* ($n = 2$), *L. booriae* ($n = 1$), and *L. fleischmannii* ($n = 1$), while one strain was closest to *L. rocourtiae* and may represent a novel species; this strain will be described in a separate presentation (Table S1; Fig. 1). Although hemolytic *L. innocua* strains were previously reported (18), all the *L. innocua* strains in our panel were nonhemolytic (data not shown).

Analysis with Genome Comparator revealed that homologs of two or more LIPI-3 genes (*lls* genes) were common in *L. innocua*, where they were detected in approximately 50% of

the isolates (Table S1; Fig. 2A and B). Outside *L. innocua*, *lIs* genes were only detected in three strains of *L. seeligeri* (Table S1; Fig. 2A and B). In most cases, LIPI-3 was truncated (partial), lacking the genes at the 3' terminus of the island that are involved in the posttranslational modification of the *lIsA*-encoded structural peptide (*lIsBYD*) and a metalloprotease gene responsible for the cleavage of the leader peptide (*lIsP*) (11, 19) (Table S1; Fig. 3). Closer scrutiny revealed two partial LIPI-3 types, i.e., *lIsAGHX* and *lIsGH* (Table S1). Interestingly, the latter was found only in three strains of *L. seeligeri* (Table S1).

To confirm the Genome Comparator data and identify *lIs* homologs in other strains, we conducted BLAST2 searches between each genome and the LIPI-3 island from *L. monocytogenes* F2365 to identify the LIPI-3 regions. When an *lIs* homolog was identified via pairwise comparisons of LIPI-3 regions but was not annotated via the annotation program Prokka, as observed with *lIsA* and *lIsP*, the genomic region where the missing *lIs* homolog is expected was rigorously annotated by finding additional, smaller open reading frames (ORFs) via getorf and conducting BLASTp searches. This analysis largely corroborated the observations from Genome Comparator but also revealed two new findings. First, *lIsX* homologs were identified in *L. seeligeri* strains in which only *lIsG* and *lIsH* were detected via Genome Comparator (Table S1; Fig. 3D); hence, these partial LIPI-3 variants are referred to here as *lIsGHX*, while those consisting of all four *lIs* genes in the 5' portion of the island are designated *lIsAGHX*. Second, an additional putative bacteriocin gene was discovered downstream of the canonical *lIsA*, which is designated *lIsA2* and further discussed below (Fig. 3).

A highly conserved full-length LIPI-3 was harbored by certain *L. innocua* strains in the same genetic location (*guaA* genomic hot spot) as in *L. monocytogenes* and included *lIsA2*. Full-length LIPI-3 was observed in just three of the 36 *L. innocua* strains: strain SKB154 (ST1086) from a black bear (*U. americanus*) and two ST1489 strains (SKWL214 and SKWL246) obtained 12 days apart from water in an urban creek (Table S1; Fig. 2A and B and 4A). Pairwise comparisons with Clustal Omega revealed that the full-length LIPI-3 was highly conserved (99.0 to 100.0% identity) among these three strains as well as among different strains of *L. monocytogenes* (99.5 to 99.9% identity) (Table S2). However, noticeably lower identity (95.5 to 97.2%) was found between the full-length LIPI-3 of the three *L. innocua* strains and *L. monocytogenes*, suggesting diversification of the island between *L. innocua* and *L. monocytogenes* (Table S2). This trend was also evidenced in the phylogenetic tree of LIPI-3 sequences, where the sequences of full-length LIPI-3 from *L. innocua* formed a separate group from their counterparts in *L. monocytogenes* (Fig. 4A).

Similar observations were obtained with the pairwise BLASTn comparisons with the LIPI-3 flanking regions (Fig. 3A). The flanking regions were highly conserved both among *L. monocytogenes* strains (>99% identity) and *L. innocua* (>98% identity) (Fig. 3A). However, a lower similarity (approximately 96%) was noted between the flanking regions of *L. monocytogenes* F2365 and *L. innocua* SKWL214 (Fig. 3A), suggesting that the LIPI-3 diversification between *L. monocytogenes* and *L. innocua* extends to the regions flanking the island as well. Nonetheless, in *L. innocua*, the full-length LIPI-3 was located next to *guaA* (Fig. 3A), a genomic hot spot that also harbors LIPI-3 and other genomic islands in *L. monocytogenes* (20, 21). However, a gene encoding a VOC family protein (*LMOF2365_RS05565* in F2365) was absent from the *L. innocua* strains with full-length LIPI-3, while it was harbored by *L. monocytogenes* strains F2365 and CFSAN023463 between *guaA* and LIPI-3 (Fig. 3A).

A multifaceted annotation strategy that employed Prokka and, when necessary, getorf and BLASTp revealed that all the full-length LIPI-3 regions, whether from *L. innocua* or from *L. monocytogenes*, contained two bacteriocin genes, i.e., *lIsA* and *lIsA2* (Fig. 3A). In *L. monocytogenes* F2365, *lIsA2* was 117 bp and overlapped *lIsA* by 73 bp (Fig. 5). BLAST2 with *lIsA* and *lIsA2* did not detect any homology between these two proteins, and a National Center for Biotechnology Information (NCBI) conserved-domain search failed to reveal any known conserved motifs. However, when analyzed with BLASTp against the NCBI nr database, the deduced *lIsA2* polypeptide exhibited 94.1% similarity with listeriolysin S family toxin, which contains the *lIsA* sequence and 15 extra

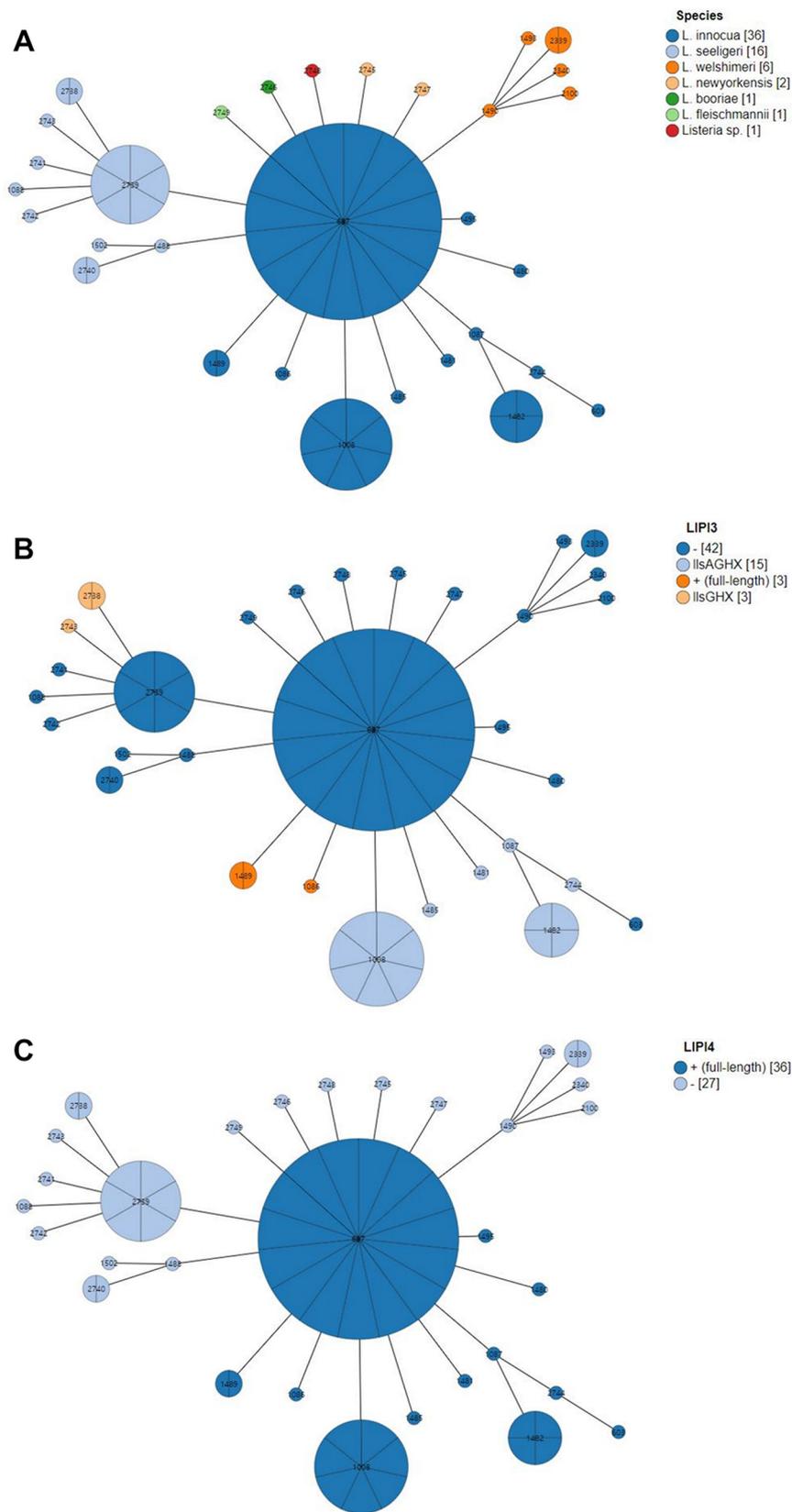


FIG 2 MSTs of the strains examined in this study. The MSTs were constructed with GrapeTree as described in Materials and Methods. Each clonal group is represented as a pie chart, which is marked with different colors representing species (A) and types of LIPI-3 (B) and LIPI-4 (C). In the color labels, the number of strains marked with each color is shown in brackets. Inside each pie chart, ST designations are included, and the size of the pie chart corresponds to the number of strains belonging to each clone.

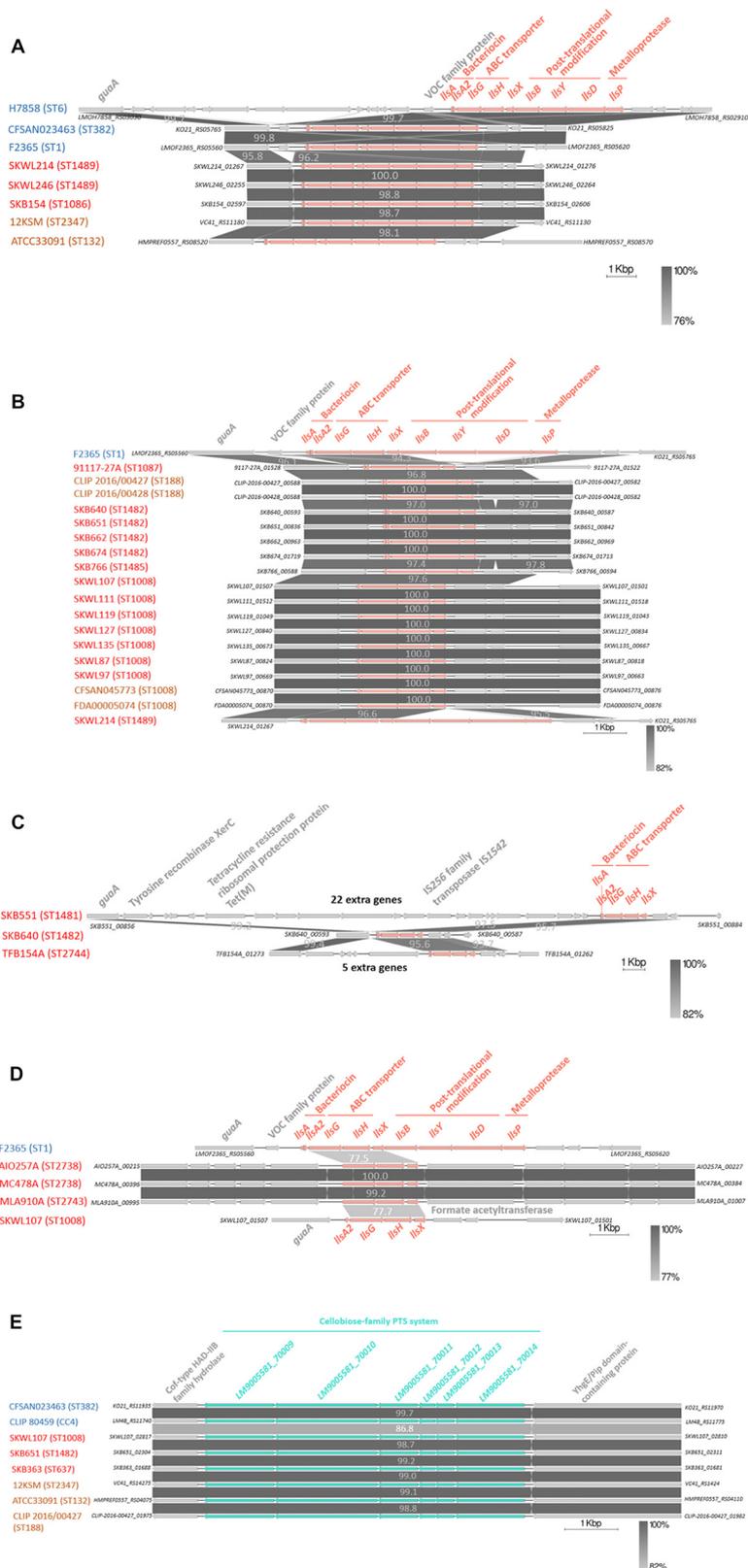


FIG 3 Comparison of LIPI-3 (A to D) and LIPI-4 (E), including flanking regions. (A) Full-length LIPI-3 regions harbored by strains from our panel (SKWL214, SKWL246, and SKB154), *L. monocytogenes* strains (H7858, CFSAN023463, and F2365), and *L. innocua* strains previously reported to harbor full-length LIPI-3 (12KSM and ATCC 33091) (18). (B) Partial LIPI-3 comprising *IISAGHX* without additional genes next to *guaA*. Strains from our panel were compared with the full-length LIPI-3 in *L. monocytogenes* F2365 and (Continued on next page)

amino acids at the C-terminal end. A promoter was detected upstream of *lIsA2*, but the Shine-Dalgarno sequence was not found via visual inspection in the vicinity of the start of *lIsA2*, possibly explaining why this gene was previously overlooked, in addition to its small size.

As with a previous study that revealed that premature stop codons (PMSCs) were often observed in *lIsA* of *L. innocua* harboring partial LIPI-3 but absent in *lIsA* in full-length LIPI-3 (18), no PMSCs were detected in *lIsA* of the *L. innocua* strains with full-length LIPI-3 in our panel (Fig. 5; Table S3). None of the other *lIs* genes in full-length LIPI-3 were found to harbor PMSCs when their lengths and protein sequence alignments were inspected (Table S3), suggesting the presence of functional LIPI-3 in these *L. innocua* strains.

The three *L. innocua* strains with full-length LIPI-3 were recently isolated from natural environments (urban water and wildlife) in North Carolina, USA, and it would be of interest to determine whether their LIPI-3 regions were similar to those in *L. innocua* strains from other sources (e.g., food processing environments), regions, and periods (17, 18). We therefore extracted and analyzed the LIPI-3 regions from two such strains (12KSM and ATCC 33091, from the food production environment in Austria in 2012 and a human case in Denmark in 1971, respectively) (18). These strains belonged to STs that were not encountered in our panel (ST2347 and ST132, respectively), suggesting that the full-length LIPI-3 is harbored by multiple clonal groups of *L. innocua* (Table S1). The LIPI-3 sequences from all five strains were found to be highly similar (97.1% to 99.9% identity), harbored the full-length LIPI-3 in the *guaA* hot spot, and included *lIsA* and *lIsA2*, with no PMSCs (Tables S2 and S3; Fig. 3A and 5). Thus, the full-length LIPI-3 in the nonpathogenic species *L. innocua* is highly conserved among strains from diverse geographical regions, sources, and periods. In the course of this analysis, we found that CFSAN045773, which was reported to harbor full-length LIPI-3 (18), actually bore a partial LIPI-3 (Fig. 3B), which is discussed below.

Partial LIPI-3 with marked genetic diversity in *lIsA* was located in the *guaA* hot spot of *L. innocua*, while genetically diversified partial LIPI-3 lacking known bacteriocin genes was harbored by *L. seeligeri* at a different genomic location. Several strains in our panel harbored truncated LIPI-3 (*lIsAGHX* and *lIsGHX*) missing several of the *lIs* genes at the 3' portion of the island (Table S1; Fig. 3B to D). The *lIsAGHX* island was common in *L. innocua*, where it was detected in 15 of the 36 strains (approximately 42%), while *lIsGHX* was found only in three strains of *L. seeligeri* from water in the Upper Snake Rock watershed (Table S1; Fig. 2A and B and 4A). The *L. innocua* strains belonged to different STs, among which ST1008 ($n = 7$) and ST1482 ($n = 4$) were predominant, and were derived from water ($n = 8$) and wildlife ($n = 7$) (Table S1; Fig. 2A and B and 4A).

FIG 3 Legend (Continued)

L. innocua SKWL214 and with the partial LIPI-3 in *L. innocua* strains previously reported by Moura et al. (CLIP 2016/00427, CLIP 2016/00428, CFSAN045773, and FDA00005074) (18). (C) Partial LIPI-3 consisting of *lIsAGHX* with extra genes between *guaA* and LIPI-3, which was harbored by strains SKB551 and TFB154A from our panel. These strains were compared with strain SKB640 harboring *lIsAGHX* without additional upstream genes. (D) Partial LIPI-3 variant consisting only of *lIsGHX* identified in *L. seeligeri* strains (AIO257A, MC478A and MLA910A). These strains were aligned with the full-length LIPI-3 from *L. monocytogenes* strain F2365 and the partial LIPI-3 *lIsAGHX* from *L. innocua* strain SKWL107. (E) LIPI-4 from representative strains from our panel, *L. monocytogenes* strains (CFSAN023463 and CLIP 80459) and *L. innocua* strains reported by Moura et al. (12KSM, ATCC 33091, and CLIP 2016/00427) (18). Representative strains were selected from our panel so that the figure does not include identical sequences and as many strains as possible are represented (Table S4). The other three *L. innocua* strains reported by Moura et al. (CLIP 2016/00428, CFSAN045773, and FDA00005074) (18) were not included in the figure, since LIPI-4 in strain CLIP 2016/00428 was identical to that of strain CLIP 2016/00427, while strains CFSAN045773 and FDA00005074 harbored the same LIPI-4 as strain SKWL107 (Table S4). Homologous regions were identified via BLAST2 conducted with the blastn algorithm and visualized with Easyfig as described in Materials and Methods. Similarity of the homologous regions is indicated in different shades of gray as marked in the gradient bar and percent sequence similarity of main homologous regions is indicated by numbers in light gray or white. Genes are signified by arrows whose direction corresponds to direction of transcription. LIPI-3 and LIPI-4 genes are in pink and turquoise, respectively, whereas those outside the islands are in light gray. LIPI-3 and LIPI-4 genes are designated near the corresponding arrows, and locus tags are provided near terminal flanking genes.

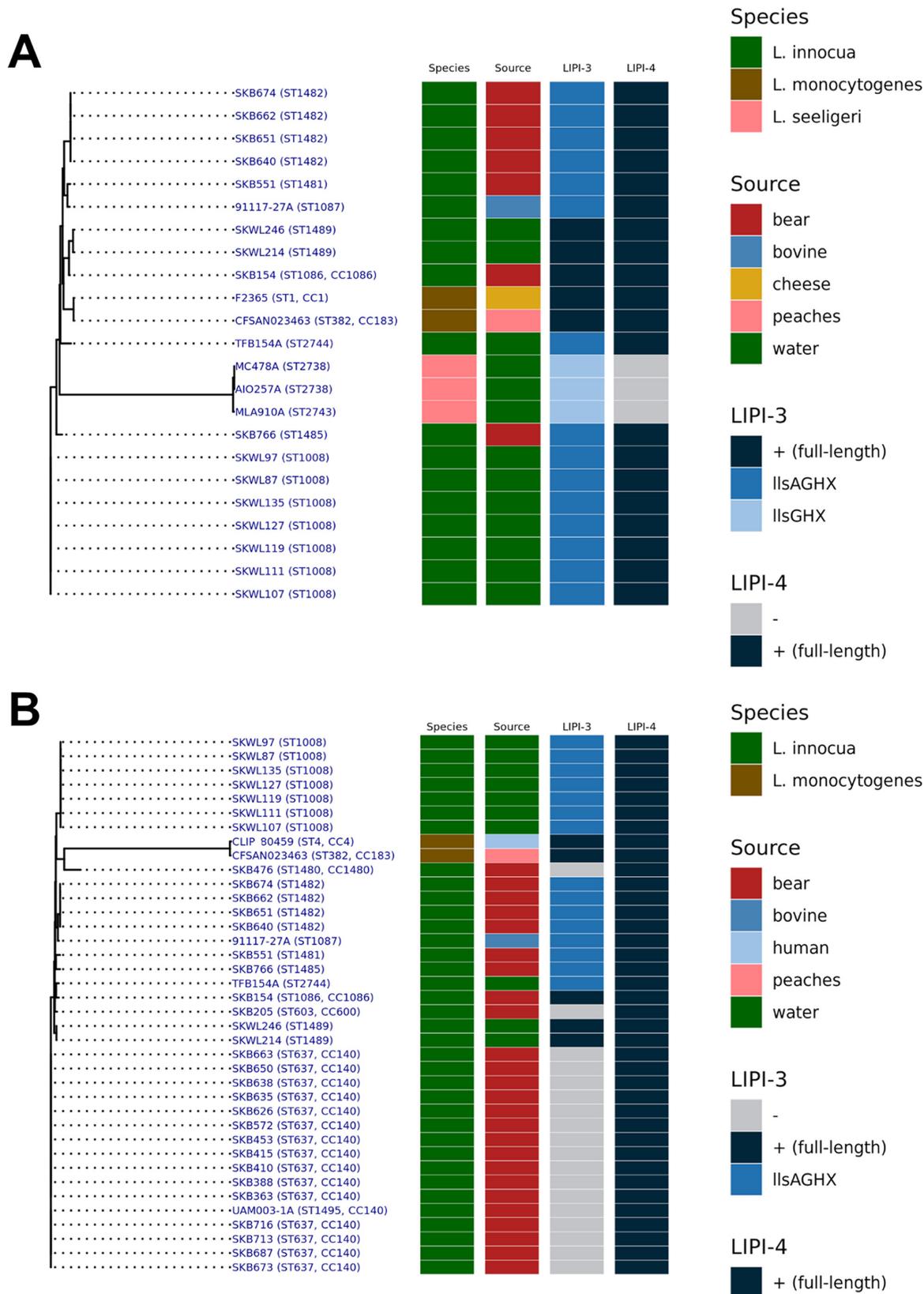


FIG 4 Phylogenetic trees of LIPI-3 (A) and LIPI-4 (B) present in the strains examined in this study and representative *L. monocytogenes* strains as described in Materials and Methods. Metadata are shown in the heat map next to the tree using different colors that represent species, source, and types of LIPI-3 and LIPI-4 (–, absence). MLST-based ST and, if available, CC designations are indicated within parentheses next to strain names.

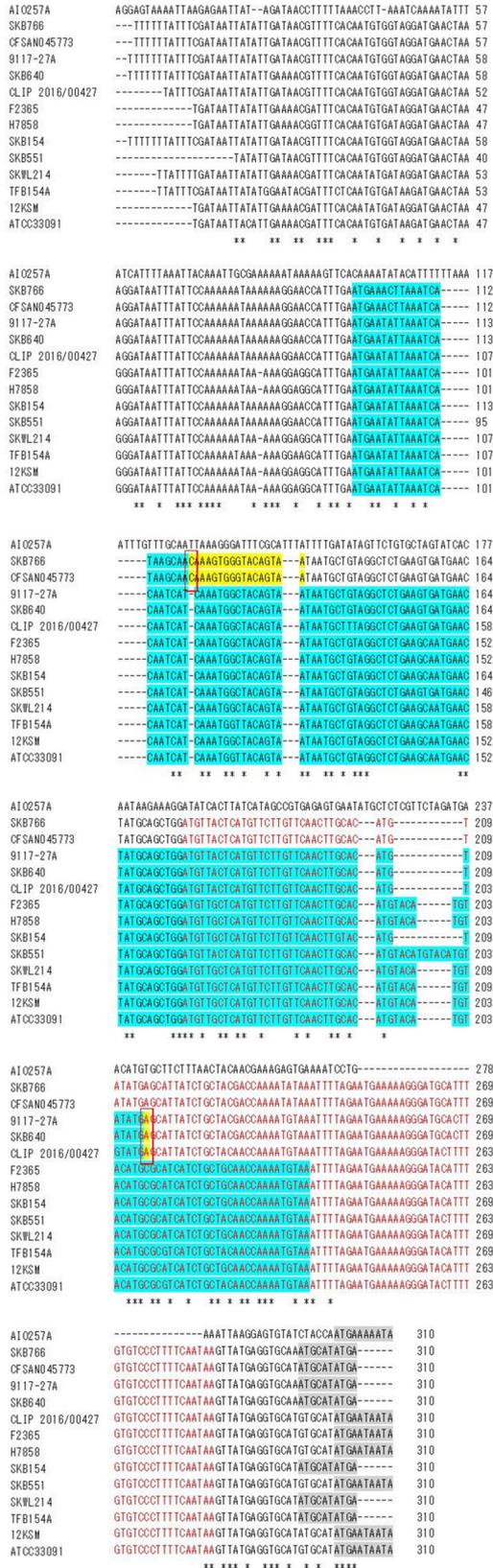


FIG 5 Alignment of the genomic region consisting of 300 bp upstream of *IISg* and 10 bp of *IISg*, which encompasses *IISa* and the novel bacteriocin gene *IISa2*. Representative sequences were chosen to remove redundant sequences. The *IISa* and *IISa2* genes are marked by blue highlighting and red font, respectively, while *IISg* gene is highlighted in gray. Changes in *IISa* due to PMSCs or SNPs are highlighted in yellow, and imperative genetic changes are marked with red rectangles.

Similarly to the full-length LIPI-3 described above, *lIsAGHX* was located in the *guaA* hot spot and exhibited conservation of the flanking regions (Fig. 3B and C). Two strains, *L. innocua* SKB551 (ST1481) and TFB154A (ST2744), harbored additional genes between *guaA* and *lIsAGHX* (Fig. 3C). *L. innocua* SKB551 harbored in this region a putative 21.6-kb DNA transposon with 22 genes, including those putatively encoding the tyrosine recombinase XerC, the tetracycline resistance ribosomal protection protein Tet(M), and the IS256 family transposase IS1542. *L. innocua* TFB154A bore a 5.3-kb sequence with five genes encoding hypothetical proteins upstream of *lIsAGHX* (Fig. 3C). These insertions provide further evidence for the *guaA* locus as a genetic hot spot prone to genetic diversification in both *L. innocua* and *L. monocytogenes*. Interestingly, the partial LIPI-3 *lIsGHX* was, as mentioned above, detected only in three strains of *L. seeligeri*, where it was inserted in an unrelated chromosomal location between a hypothetical protein gene and a formate acetyltransferase gene (Fig. 3D). The extent to which this location may serve as a genetic hot spot in *L. seeligeri* remains to be determined.

Seven of the 15 *L. innocua* strains with the partial LIPI-3 *lIsAGHX* harbored both *lIsA* and *lIsA2*, as also observed in the full-length LIPI-3 of *L. innocua* and *L. monocytogenes* (Fig. 3B and C). However, five of these seven strains shared a truncated *lIsA* of 117 bp compared with 150-bp *lIsA* identified in full-length LIPI-3, due to a common C-to-A substitution that resulted in a PMSC (Fig. 5; Table S3). The remaining eight *lIsAGHX*-harboring strains harbored *lIsA2* but lacked the intact *lIsA* due to an insertion of a cytosine immediately after the start of *lIsA*, which led to a frameshift and PMSCs, eventually resulting in the failure to annotate this gene (Fig. 3B and 5). These genetic changes, i.e., PMSCs and frameshifts, appear to be limited to *lIsA*, since such variations were not observed when the length of other *lIs* genes was analyzed and protein sequences encoded by other *lIs* genes were aligned (Table S3; data not shown).

In pairwise comparisons with Clustal Omega, *lIsAGHX*, which is expected to yield both *lIsA* and *lIsA2*, exhibited similarities ranging from 95.3% to 100.0% while *lIsAGHX*, which encodes *lIsA2* but not *lIsA*, was more highly conserved (97.5% to 100.0% identity) (Table S2). Compared with the full-length LIPI-3 regions from *L. innocua* and *L. monocytogenes*, *lIsAGHX* displayed higher similarities with the full-length LIPI-3 of *L. innocua* than with LIPI-3 of *L. monocytogenes* (Table S2). These tendencies were also observed in the pairwise comparisons employing BLASTn for LIPI-3 and its flanking regions (Fig. 3B).

Analysis of several *L. innocua* strains with partial LIPI-3 reported by Moura et al. (CLIP 2016/00427, CLIP 2016/00428, and FDA00005074) as well as CFSAN045773 (18) revealed that all harbored the partial LIPI-3 *lIsAGHX* downstream of *guaA* (Fig. 3B). The partial LIPI-3 regions from strains CLIP 2016/00427 and CLIP 2016/00428, which were almost identical (100.0% identity with only a few mismatches), exhibited lower similarity (95.4 to 96.8% identity) to *lIsAGHX* from our panel (Table S2). However, the presence of both *lIsA* and *lIsA2* and the similarity in the DNA sequence including flanking regions suggest that the partial LIPI-3 of *L. innocua* CLIP 2016/00427 and CLIP 2016/00428 closely resembles those of ST1482 strains in our panel (Fig. 3B). Furthermore, in these strains the PMSC in *lIsA* was at the same location as in ST1482 (Fig. 5; Table S3). On the other hand, the *L. innocua* ST1008 strains FDA00005074 and CFSAN045773 harbored a partial LIPI-3 identical to our strains of ST1008 and also encoded only the intact *lIsA2* but not *lIsA* in LIPI-3 (Table S2; Fig. 3B), providing evidence for clonal conservation of the partial LIPI-3.

Analysis of the *L. seeligeri lIsGHX* revealed that it was highly conserved (99.7% to 100.0% identity) in this species but differed considerably from other LIPI-3 in *L. innocua* or *L. monocytogenes*, whether partial or full length, with similarities ranging from 72.2% to 73.5% in pairwise comparisons using Clustal Omega (Table S2). This diversification of the *L. seeligeri lIsGHX* was also manifested in the phylogenetic tree of LIPI-3, where *lIsGHX* formed a group distinct from other LIPI-3 cassettes (Fig. 4A), as well as in pairwise comparisons of LIPI-3 and its flanking genes with *L. monocytogenes* F2365 (harboring full-length LIPI-3) and *L. innocua* SKWL107 (harboring a partial LIPI-3) (Fig. 3D). In spite of various attempts, we could not identify any genes homologous to known bacteriocin genes near *lIsGHX*

(Fig. 3D); however, we cannot exclude the possibility that an unknown bacteriocin might be encoded in the vicinity of this partial LIPI-3.

LIPI-3 is undergoing genetic diversification in bacteriocin genes after being horizontally acquired by nonpathogenic *Listeria* spp. Our data and those from previous studies (17, 18) indicate that, when present in *L. innocua*, LIPI-3 was integrated in the same locus as in *L. monocytogenes* but was frequently truncated, missing the four genes in the 3' portion of the island, and *lIsA* can harbor frequent PMSCs. Our analysis furthermore revealed a frameshift in *lIsA* caused by the insertion of one nucleotide in several *L. innocua* strains and also identified *L. seeligeri* strains with a partial LIPI-3 lacking bacteriocin genes (Fig. 3D and 5). Such findings suggest that LIPI-3 might be undergoing reductive evolution, resulting in the degeneration of this island. However, the commonly missing genes were all involved in the modification of *lIsA*, and it is conceivable that their functions are performed by genes involved in processing other bacteriocins, potentially resulting in the production of an intact or differently modified bacteriocin.

With the exception of *lIsA* and *lIsA2*, the GC content of *lIs* genes was lower than the genome average of *L. innocua* (37%) (22), ranging from 24.3 to 30.7% (Table S3), suggesting that they were acquired via horizontal gene transfer (HGT). On the other hand, the GC content of *lIsA* ranged between 35.0 and 37.3% in strains harboring a partial LIPI-3 and between 37.3 and 38.0% when *lIsA* was on a full-length LIPI-3 (Table S3). The newly identified *lIsA2* had GC content from 31.5 to 35.0%, which tended to be higher when *lIsA2* was located next to the intact *lIsA* than when *lIsA2* resided along with a PMSC in the otherwise complete *lIsA* sequence, leading to the absence of *lIsA*, in LIPI-3 (Table S3). As with *lIsA*, the highest GC content was observed in *lIsA2* genes harbored by full-length LIPI-3 regions (34.2 to 35.0%) (Table S3). Variability in the GC content of these bacteriocin genes was not surprising considering their small size of less than 160 bp, where GC content can be affected even by a small number of nucleotide changes. Our data suggest that *lIsA* or *lIsA2* may have been acquired independently from other LIPI-3 genes and may be undergoing genetic changes that result in lower GC content, in agreement with the trend toward diversification of LIPI-3 genes.

Previous studies showed that in *L. monocytogenes*, the *lIsA*-encoded listeriolysin S is expressed only in the animal intestine and is a contact-mediated bacteriocin that impacts virulence by modulating the composition of the gut microbiota (12, 13, 23). It is therefore perplexing that LIPI-3 is harbored by many strains of *L. innocua* and occasional strains of *L. seeligeri*. LIPI-3, whether truncated or full length, in nonpathogenic *Listeria* spp. did not lead to hemolysis, since none of *L. innocua* strains in our panel, including those with the full-length LIPI-3, exhibited hemolysis (data not shown). This was also supported by the study conducted by Moura et al. in 2019, which reported that hemolysis observed in *L. innocua* is closely associated with the harborage of the full-length LIPI-1 but not with the presence of LIPI-3 (18). The potential roles of LIPI-3 in the ecology and adaptations of these nonpathogenic strains remain to be elucidated. It is conceivable that LIPI-3 may be an obsolete genetic element in several of these nonpathogenic strains that harbor truncated versions of the island and *lIsA* with PMSCs and frameshifts. On the other hand, our findings clearly indicate the presence of full-length and likely functional LIPI-3 in *L. innocua* strains from wildlife and surface water. It is thus plausible that LIPI-3 in such nonpathogenic strains may confer specific fitness advantages by modulating the microbial community in the corresponding environmental niches (e.g., gut or feces of the wildlife host), similarly to its roles observed in *L. monocytogenes*.

Intact LIPI-4 was stably maintained in all *L. innocua* strains, in the same location as in *L. monocytogenes*. As discussed above, the 6-gene pathogenicity island LIPI-4 was originally discovered during the analysis of genomic regions unique to the *L. monocytogenes* hypervirulent clone CC4 and was shown to be critical for neurovirulence in a murine model (15). Subsequent findings revealed that LIPI-4 was also harbored by other clones of *L. monocytogenes* (10, 14, 16), and a study of LIPI-3 in *L. innocua* reported that the investigated strains also harbored LIPI-4 (18). However, there has been overall surprisingly limited discussion of the presence, diversity, and roles of this pathogenicity island in nonpathogenic *Listeria* spp. Initial analysis of the strains in our panel with Genome Comparator

revealed that full-length LIPI-4 was harbored only by *L. innocua*, where it was detected in all 36 strains of this species regardless of their origin and MLST-derived genotype (Table S1; Fig. 2A and C). Partial LIPI-4 cassettes were never encountered, and no PMSCs were identified in any of the LIPI-4 genes (Tables S1 and S5), in agreement with a previous report on *L. innocua* strains from other sources (18). In all the *L. innocua* strains in our panel as well as those previously examined (18), LIPI-4 was in the same genetic location as in *L. monocytogenes*, between a Cof-type HAD-IIB family hydrolase gene (*Imo2359* homolog) and a gene encoding a YhgE/Pip domain-containing protein (*Imo2360* homolog), which has not been previously identified among genetic hot spots in *L. monocytogenes* (Fig. 3E) (20).

LIPI-4 was highly conserved (97.0% to 100% identity) among the *L. innocua* strains from our panel as well as with those investigated previously (18) (Table S4). LIPI-4 harbored by different clones of *L. monocytogenes*, i.e., the ST4 strain CLIP 80459 and the ST382 strain CFSAN023463, was also found to have similarly high conservation (99.9% identity) (Table S4). However, noticeably lower similarity (83.7 to 84.0% identity) was noted between LIPI-4 of *L. innocua* and its counterpart in *L. monocytogenes* (Table S4), suggesting that *L. innocua* and *L. monocytogenes* harbor two distinct LIPI-4 variants. This finding was clearly visualized in the phylogenetic tree of LIPI-4 sequences (Fig. 4B) and BLASTn analysis of the DNA sequences encompassing LIPI-4 and adjacent genes (Fig. 3E). The overall GC content of LIPI-4 was below the genome average of *L. innocua* (37%) and *L. monocytogenes* (39%) (22) (Table S5), suggesting that LIPI-4 genes were acquired via HGT.

The ubiquity of full-length LIPI-4 in *L. innocua* and the observed absence of PMSCs are in marked contrast with the findings regarding LIPI-3, which was harbored only by certain strains, was often partial with several genes missing, and exhibited PMSCs and frameshifts, suggesting a lack of functionality. The ubiquity and apparent genetic stability of LIPI-4 in *L. innocua* may be indicative of specific selection pressures to which this species is exposed that necessitate preservation and functionality of this island. Our findings support the hypothesis that LIPI-4 was originally harbored by the ancestral lineage that subsequently differentiated into the pathogen *L. monocytogenes* and the nonpathogenic species *L. innocua* and that it was widely retained in the latter but only in clonal complex 4 (CC4) and a few other clones of the former, with subsequent differentiation reflected in the observed sequence divergence between the LIPI-4 variants in the two species. Alternatively, LIPI-4 might follow a similar evolutionary route toward degeneration in *L. innocua* as observed in LIPI-3, but it might not have been given sufficient time to amass ample genetic changes. However, this hypothesis cannot explain the fact that LIPI-4 is encountered among only a few clones of *L. monocytogenes*, which suggests that, in the same time span, LIPI-4 in *L. monocytogenes* might have experienced such drastic genetic changes that this island was lost from a majority of the *L. monocytogenes* population.

It is noteworthy that in *L. monocytogenes*, both LIPI-3 and LIPI-4 are harbored primarily by certain clones of lineage I, mostly of serotype 4b, and are extremely rare in other lineages (11, 14–16). This distribution is suggestive of a common habitat shared by lineage I *L. monocytogenes* and *L. innocua*, with microbial community composition and environmental conditions promoting maintenance of the pathogenicity islands. The similar teichoic acid composition between serotype 4b *L. monocytogenes* and *L. innocua* (24) may promote HGT via transducing phages utilizing specific teichoic acid receptors and may contribute to the observed distribution of these islands. Further studies are needed to characterize the roles of the islands in the environmental ecology and adaptations of nonpathogenic *Listeria* spp. such as *L. innocua*. This will be especially valuable for LIPI-4, which may constitute an intriguing model for genetic elements evolving in response to differential selection pressures in pathogenic versus nonpathogenic species.

Conclusions. In this study, we scrutinized the genetic diversity and evolution of two pathogenicity islands, LIPI-3 and LIPI-4, that enhance the virulence of the human pathogen *L. monocytogenes* but were also encountered in nonpathogenic *Listeria* spp. While both islands appear to be acquired via HGT, LIPI-3 showed various signs of

genetic diversification such as the absence of several genes and PMSCs in the bacteriocin gene *lIsA*, whereas for LIPI-4, a single variant distinct from the *L. monocytogenes* counterpart was stably harbored by all tested *L. innocua* strains but in no other nonpathogenic *Listeria* spp. These findings demonstrate divergent fates that a horizontally transferred pathogenicity island faces within nonpathogenic bacterial species, possibly due to dissimilar selection pressures to maintain such an island during the life cycle of nonpathogenic bacteria. Investigation of the functions of LIPI-3 and LIPI-4 in nonpathogenic *Listeria* spp. would be a fascinating topic for future research to understand how genes evolve and play distinct roles in a different genomic background.

MATERIALS AND METHODS

Bacterial strains and growth conditions. The nonpathogenic *Listeria* species strains investigated in this study are listed in Table S1. They included seven nonpathogenic *Listeria* spp. and, with the exception of one strain from bovine feces, were isolated in the course of previous studies from surface water, water-immersed rocks or debris, and wildlife (American black bears [*U. americanus*] (Table S1) using previously described methods (25, 26). The strains were routinely cultured at 37°C in tryptic soy broth with 0.6% yeast extract (Becton, Dickinson and Co., Sparks, MD, USA) or in brain heart infusion (BHI) broth (Becton, Dickinson and Co.) and were preserved at –80°C in BHI broth with 20% glycerol.

WGS analysis. Whole-genome sequences were obtained as described previously (25). Briefly, genomic DNA was extracted with a DNeasy blood and tissue kit (Qiagen, Valencia, CA, USA) from overnight cultures grown in BHI broth and then was utilized to construct libraries with a Nextera XT DNA library preparation kit (Illumina, San Diego, CA, USA). Genome sequence was obtained with a NextSeq 500/550 high-output kit v2.5 and a NextSeq 500 sequencer (300 cycles, 2 × 150 bp) (Illumina). Resultant FASTQ files were quality-trimmed and *de novo* assembled with CLC Genomics Workbench v. 20 (Qiagen, Aarhus, Denmark) and the quality of assemblies was measured with QUAST v.4.6.4 (27). *Listeria* sp. designations were determined employing the pyani software package (<https://github.com/widdowquinn/pyani>) (28). MLST-based ST and CC designations were assigned employing the MLST package (<https://github.com/tseemann/mlst>) as previously described (25) or the PubMLST database (<https://bigsd.b.pasteur.fr/listeria/>) maintained by the Institut Pasteur. CC designations were double-checked with the PubMLST database. Genes were annotated with Prokka (version 1.14.5; available in the Docker image downloaded at <https://hub.docker.com/r/staphb/prokka>) (29).

Retrieval of genome sequences from the NCBI. Selected genome sequences of other *Listeria* strains were retrieved from the NCBI database. To obtain the prototype LIPI-3, we downloaded from NCBI the genome of *L. monocytogenes* F2365, ST1, accession no. [NC_002973.6](https://ncbi.nlm.nih.gov/assembly/NC_002973.6) (11, 30, 31). For LIPI-4, we downloaded the genome of the ST4 strain *L. monocytogenes* CLIP 80459 (accession no. [NC_012488.1](https://ncbi.nlm.nih.gov/assembly/NC_012488.1)) (31). Additional *L. monocytogenes* genomes from NCBI included the 2014 stone fruit outbreak strain CFSAN023463, ST382, harboring both LIPI-3 and LIPI-4 (accession no. [NZ_CP012021.1](https://ncbi.nlm.nih.gov/assembly/NZ_CP012021.1)) (16), and the ST6 strain H7858 from the 1998–1999 hot dog outbreak, harboring LIPI-3 but lacking LIPI-4 (accession no. [GCF_000167155.1](https://ncbi.nlm.nih.gov/assembly/GCF_000167155.1)) (11, 30). To compare our strains with LIPI-3 and LIPI-4 regions previously reported in *L. innocua*, we included the six *L. innocua* strains ATCC 33091, CFSAN045773, and 12KSM (reported to harbor both LIPI-4 and full-length LIPI-3; belonging to ST132, ST1008, and ST2347, respectively; accession no. [GCF_000241405.1](https://ncbi.nlm.nih.gov/assembly/GCF_000241405.1), [SRR3345521](https://ncbi.nlm.nih.gov/assembly/SRR3345521), and [GCF_000960735.1](https://ncbi.nlm.nih.gov/assembly/GCF_000960735.1), respectively); CLIP 2016/00427 and CLIP 2016/00428 (reported to harbor LIPI-4 and a partial LIPI-3 with a premature stop codon in *lIsA*; ST188; accession no. [ERR2729679](https://ncbi.nlm.nih.gov/assembly/ERR2729679) and [ERR2729680](https://ncbi.nlm.nih.gov/assembly/ERR2729680), respectively); and FDA00005074 (reported to harbor LIPI-4 and a partial LIPI-3 without *lIsA*; ST1008; accession no. [SRR3137591](https://ncbi.nlm.nih.gov/assembly/SRR3137591)) (18). GenBank and FASTA files of complete genome sequences (F2365, CLIP 80459, and CFSAN023463) were downloaded directly from the NCBI website, while those from the incomplete genomes (H7858, ATCC 33091, and 12KSM) were downloaded programmatically from their FTP sites using custom Bash scripts. For genomes of CFSAN045773, CLIP 2016/00427, CLIP 2016/00428, and FDA00005074, FASTQ files were downloaded from the NCBI SRA database and assembled with SPAdes at PATRIC (<https://www.patricbrc.org/>), generating FASTA files containing contigs which were subsequently annotated with Prokka as described above (29, 32, 33). ST designations of the previously studied *L. innocua* strains were determined with PubMLST database.

Extraction of LIPI-3 and LIPI-4 from the genome sequence data. LIPI-3 and LIPI-4 genes in our strain panel were initially examined with Genome Comparator (https://bigsd.b.pasteur.fr/cgi-bin/bigsd/bigsd.pl?db=pubmlst_listeria_isolates&page=plugin&name=GenomeComparator) using default parameters with the Virulence scheme, which indicated the presence or absence of each gene and its allele number. Each GenBank file was visualized as a circular map with CGView Server (<http://cgview.ca/>) (34), and regions homologous to the sequence encompassing the prototype LIPI-3 or LIPI-4 along with its flanking regions were marked on the map. The query sequence for LIPI-3 BLAST2 contained genes from *LMOF2365_RS05560* (*guaA*) to *LMOF2365_RS05620* in *L. monocytogenes* F2365, whereas that for LIPI-4 BLAST2 consisted of genes from *LM4B_RS11740* to *LM4B_RS11775* in *L. monocytogenes* CLIP 80459. These sequences were determined using previous studies and information available in the Virulence Factor Database (<http://www.mgc.ac.cn/cgi-bin/VFs/pai.cgi?PAIID=P025> for LIPI-3 and <http://www.mgc.ac.cn/cgi-bin/VFs/pai.cgi?PAIID=P026> for LIPI-4) (11, 31, 35). Regions homologous to the query sequence in each genome were manually inspected to identify a region including LIPI-3 or LIPI-4 and adjacent genes,

which was extracted from the GenBank file of the genome with custom scripts using Bash and Biopython, generating both GenBank and FASTA files that were utilized in subsequent analyses (36).

Identification of additional genes in LIPI-3. Preliminary comparison of LIPI-3 and its adjacent genes in F2365 with the counterpart in several Prokka-annotated genomes revealed that *lIsA* and *lIsP* were overlooked by Prokka. Hence, we utilized the ORF detection program getorf (version EMBOSS:6.5.7.0) in the EMBOSS Docker image (version v6.5.7_cv2 [<https://hub.docker.com/r/biocontainers/emboss>]) with the parameters modified so that any ORFs ranging from 100 bp to 4,000 bp could be identified in both strands within a designated region, and protein sequences encoded in the ORFs translated with bacterial genetic code were obtained (-table 11 -minsize 100 -maxsize 4000 -find 1 -reverse "Y" -flanking 0) (37, 38). Searches for *lIsA* were done within the intergenic region upstream of *lIsG* and its 101-bp flanking regions in all the LIPI-3 regions, whereas *lIsP* was looked for in the intergenic region downstream of *lIsD* and its 101-bp flanking regions in the full-length LIPI-3 regions. The protein sequences of the ORFs identified via getorf were subjected to a BLASTp search against the NCBI nr database using Biopython, and only the ORFs whose BLASTp hits contain designated keywords ("listeriolysin" for *lIsA* and "CPBP" for *lIsP*) were incorporated into the GenBank file containing LIPI-3 and flanking regions (36, 39). When no listeriolysin or bacteriocin gene was identified in the vicinity of *lIsG* (strains AIO257A, MC478A, and MLA910A), the getorf program was implemented again to find smaller ORFs (30 to 100 bp), and their BLASTp outputs were manually inspected. Also, five genes flanking LIPI-3 in each side in strain AIO257A were extracted with Biopython and bioawk (<https://github.com/lh3/bioawk>) and further analyzed with BLASTp in an attempt to identify potential bacteriocin genes near LIPI-3 (36, 39). When *lIsA* and *lIsP* were already annotated in the genome, as observed for strains F2365, CFSAN023463, H7858, ATCC 33091, and 12KSM, we searched for ORFs of 100 to 4,000 bp to identify any additional bacteriocin genes in the intergenic region upstream of *lIsA* and its 101-bp flanking regions, as well as the intergenic region upstream of *lIsG* and its 101-bp flanking regions. The resulting ORFs were analyzed with BLASTp, and those similar to listeriolysin or bacteriocin were added to the GenBank files that include LIPI-3 and its flanking regions. Subsequent analyses were conducted with the GenBank files of the LIPI-3 region containing flanking genes that were obtained with these annotation approaches.

To investigate the characteristics of the newly identified *lIsA2*, *lIsA* and *lIsA2* in *L. monocytogenes* F2365 were compared with BLAST2 using the blastp algorithm, and *lIsA2* was also subjected to a BLASTp search against the NCBI nr database and an NCBI conserved-domain search (39, 40). Promoters were searched for in the 200-bp region upstream of the start codon of F2365 *lIsA2* using BPROM (41), and the Shine-Dalgarno sequence (42) was visually inspected.

Detection of PMSCs and evidence of HGT in LIPI-3 and LIPI-4 genes. To identify PMSCs in LIPI-3 and LIPI-4 genes, the DNA and protein sequences of homologous genes were extracted using Biopython from the GenBank files (36). PMSCs were detected by comparing the size of DNA sequences of homologous genes and aligning the protein sequences via Clustal Omega (version 1.2.1) in the clustal-omega Docker image of the BioContainers project (version v1.2.1_cv5 [<https://hub.docker.com/r/biocontainers/clustal-omega>]) (38, 43), since those with PMSCs are expected to be shorter than their homologs and exhibit similarity only in a part of the protein sequence containing the N terminus of the intact protein.

To identify single nucleotide polymorphisms (SNPs) in *lIsA* that lead to significant genetic changes, including PMSCs, DNA sequences spanning 300 bp upstream region of *lIsG* and 10 bp of *lIsG* were extracted with bioawk from the FASTA files of LIPI-3 and its flanking regions, so that all the bacteriocin genes in LIPI-3 were included. These sequences were aligned with Clustal Omega in the clustal-omega Docker image of the BioContainers project (38, 43). Then, the output was visually inspected to identify PMSCs or SNPs that may result in frameshifts. The GC content of individual genes was determined using Biopython to obtain evidence of HGT (36).

Construction of phylogenetic trees and MSTs. Pan-genome SNPs were aligned with kSNP (version 3.1; available in the Docker image obtained at <https://hub.docker.com/r/staphb/k SNP3>) generating the alignment file in FASTA format, which was utilized to construct a phylogenetic tree with MEGA X (available in the Docker image downloaded at https://hub.docker.com/r/pegi3s/megax_cc/) using the neighbor-joining method with 1,000 bootstraps (44, 45). The Newick tree file generated with MEGA X was visualized along with the strain metadata using R library ggtree (version v3.0.1; available in the Docker image of version 3.0.1-r41hdfd78af_0 obtained at <https://quay.io/repository/biocontainers/bioconductor-ggtree?tab=tags>) (46). Phylogenetic trees were also constructed with DNA sequences containing only LIPI-3 or LIPI-4 genes. FASTA files of the LIPI-3 and LIPI-4 regions without any flanking sequences were generated with custom scripts using Bash and Biopython from the GenBank files containing LIPI-3 and its flanking regions (36). These sequences were first aligned with Clustal Omega provided by EMBL-EBI (<https://www.ebi.ac.uk/Tools/msa/clustalo/>), resulting in the alignment file in FASTA format, from which a phylogenetic tree was constructed with MEGA X and visualized with ggtree as described for the genome phylogenetic tree (43, 45, 46). Relatedness among the genomes and correlations with metadata were visualized in minimum spanning trees (MSTs) generated with GrapeTree (<https://github.com/achtman-lab/GrapeTree/releases>) after retrieving the MLST allele information of housekeeping genes corresponding to the ST designations from the PubMLST database (47).

Pairwise comparisons of LIPI-3 and LIPI-4. To compare LIPI-3 regions, FASTA files of the LIPI-3 regions without any flanking sequences were aligned with Clustal Omega in the clustal-omega Docker image of the BioContainers project (38, 43). The percent similarity of each pair was identified by inspecting the percent identity matrix file generated by this analysis. Additionally, to compare the similarity of LIPI-3 sequence and its location, LIPI-3 regions along with their flanking regions were compared pairwise among the strains in our panel and with those from *L. monocytogenes* strains H7858, F2365, and CFSAN023463 and from the previously reported *L. innocua* strains ATCC 33091, CFSAN045773, 12KSM, CLIP 2016/00427, CLIP 2016/00428, and FDA00005074 (18). This analysis was conducted with BLAST2

(version 2.10.0+; included in the Docker image downloaded at <https://hub.docker.com/r/ncbi/blast>) using the blastn algorithm and was visualized with Easyfig (39, 48). These processes were also applied to LIPI-4 with the exception that *L. monocytogenes* strains CFSAN023463 and CLIP 80459 were utilized in the comparisons.

Data availability. The raw sequence reads were deposited in the NCBI database, and the accession numbers are listed in Table S1.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, XLSX file, 0.02 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.02 MB.

SUPPLEMENTAL FILE 3, XLSX file, 0.01 MB.

SUPPLEMENTAL FILE 4, XLSX file, 0.02 MB.

SUPPLEMENTAL FILE 5, XLSX file, 0.01 MB.

ACKNOWLEDGMENTS

This work was partially funded by USDA AFRI-ELI under award 2017-67012-26001, USDA National Institute of Food and Agriculture (USDA-NIFA) under award 2019-67017-29581, and the National Research Foundation of Korea (NRF-2020R1F1A1064262). Any opinions, findings, conclusions or recommendations expressed in this study are those of the authors and do not necessarily reflect the view of the USDA or other funders.

We thank James Jackson and Jeffrey Niedermeyer for providing technical assistance and isolating strains, respectively. Also, we are grateful to the team of curators of the Institut Pasteur MLST system (Paris, France [<http://bigsd.bpasteur.fr/>]), who imported novel alleles, profiles, and/or isolates.

REFERENCES

- Painter J, Slutsker L. 2007. Listeriosis in humans, p 85–110. In Ryser ET, Marth EH (ed), *Listeria*, listeriosis, and food safety, 3rd ed. CRC Press, Boca Raton, FL.
- Schuchat A, Swaminathan B, Broome CV. 1991. Epidemiology of human listeriosis. *Clin Microbiol Rev* 4:169–183. <https://doi.org/10.1128/CMR.4.2.169>.
- Orsi RH, den Bakker HC, Wiedmann M. 2011. *Listeria monocytogenes* lineages: genomics, evolution, ecology, and phenotypic characteristics. *Int J Med Microbiol* 301:79–96. <https://doi.org/10.1016/j.ijmm.2010.05.002>.
- Swaminathan B, Gerner-Smith P. 2007. The epidemiology of human listeriosis. *Microbes Infect* 9:1236–1243. <https://doi.org/10.1016/j.micinf.2007.05.011>.
- Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson M-A, Roy SL, Jones JL, Griffin PM. 2011. Foodborne illness acquired in the United States—major pathogens. *Emerg Infect Dis* 17:7–15. <https://doi.org/10.3201/eid1701.p11101>.
- de Noordhout CM, Devleeschauwer B, Angulo FJ, Verbeke G, Haagsma J, Kirk M, Havelaar A, Speybroeck N. 2014. The global burden of listeriosis: a systematic review and meta-analysis. *Lancet Infect Dis* 14:1073–1082. [https://doi.org/10.1016/S1473-3099\(14\)70870-9](https://doi.org/10.1016/S1473-3099(14)70870-9).
- Schlech WF. 2019. Epidemiology and clinical manifestations of *Listeria monocytogenes* infection. *Microbiol Spectr* 7:GPP3-0014-2018. <https://doi.org/10.1128/microbiolspec.GPP3-0014-2018>.
- Kathariou S. 2002. *Listeria monocytogenes* virulence and pathogenicity, a food safety perspective. *J Food Prot* 65:1811–1829. <https://doi.org/10.4315/0362-028x-65.11.1811>.
- Vázquez-Boland JA, Kuhn M, Berche P, Chakraborty T, Domínguez-Bernal G, Goebel W, González-Zorn B, Wehland J, Kreft J. 2001. *Listeria* pathogenesis and molecular virulence determinants. *Clin Microbiol Rev* 14:584–640. <https://doi.org/10.1128/CMR.14.3.584-640.2001>.
- Disson O, Moura A, Lecuit M. 2021. Making sense of the biodiversity and virulence of *Listeria monocytogenes*. *Trends Microbiol* 29:811–822. <https://doi.org/10.1016/j.tim.2021.01.008>.
- Cotter PD, Draper LA, Lawton EM, Daly KM, Groeger DS, Casey PG, Ross RP, Hill C. 2008. Listeriolysin S, a novel peptide haemolysin associated with a subset of lineage I *Listeria monocytogenes*. *PLoS Pathog* 4:e1000144. <https://doi.org/10.1371/journal.ppat.1000144>.
- Quereda JJ, Dussurget O, Nahori M-A, Ghazlane A, Volant S, Dillies M-A, Regnault B, Kennedy S, Mondot S, Villoing B, Cossart P, Pizarro-Cerda J. 2016. Bacteriocin from epidemic *Listeria* strains alters the host intestinal microbiota to favor infection. *Proc Natl Acad Sci U S A* 113:5706–5711. <https://doi.org/10.1073/pnas.1523899113>.
- Quereda JJ, Nahori MA, Meza-Torres J, Sachse M, Titos-Jiménez P, Gomez-Laguna J, Dussurget O, Cossart P, Pizarro-Cerdá J. 2017. Listeriolysin S is a streptolysin S-like virulence factor that targets exclusively prokaryotic cells in vivo. *mBio* 8:e00259-17. <https://doi.org/10.1128/mBio.00259-17>.
- Moura A, Criscuolo A, Pousee H, Maury MM, Leclercq A, Tarr C, Björkman JT, Dallman T, Reimer A, Enouf V, Larssonneur E, Carleton H, Bracq-Dieye H, Katz LS, Jones L, Touchon M, Tourdjman M, Walker M, Stroika S, Cantinelli T, Chenal-Francoise V, Kucerova Z, Rocha EPC, Nadon C, Grant K, Nielsen EM, Pot B, Gerner-Smith P, Lecuit M, Brisse S. 2016. Whole genome-based population biology and epidemiological surveillance of *Listeria monocytogenes*. *Nat Microbiol* 2:16185. <https://doi.org/10.1038/nmicrobiol.2016.185>.
- Maury MM, Tsai Y-H, Charlier C, Touchon M, Chenal-Francoise V, Leclercq A, Criscuolo A, Gaultier C, Roussel S, Brisabois A, Disson O, Rocha EPC, Brisse S, Lecuit M. 2016. Uncovering *Listeria monocytogenes* hypervirulence by harnessing its biodiversity. *Nat Genet* 48:308–313. <https://doi.org/10.1038/ng.3501>.
- Chen Y, Luo Y, Pettengill J, Timme R, Melka D, Doyle M, Jackson A, Parish M, Hammack TS, Allard MW, Brown EW, Strain EA. 2017. Singleton sequence type 382, an emerging clonal group of *Listeria monocytogenes* associated with three multistate outbreaks linked to contaminated stone fruit, caramel apples, and leafy green salad. *J Clin Microbiol* 55:931–941. <https://doi.org/10.1128/JCM.02140-16>.
- Clayton EM, Daly KM, Guinane CM, Hill C, Cotter PD, Ross PR. 2014. Atypical *Listeria innocua* strains possess an intact LIPI-3. *BMC Microbiol* 14:58. <https://doi.org/10.1186/1471-2180-14-58>.
- Moura A, Disson O, Lavina M, Thouvenot P, Huang L, Leclercq A, Fredriksson-Ahomaa M, Eshwar AK, Stephan R, Lecuit M. 2019. Atypical hemolytic *Listeria innocua* isolates are virulent, albeit less than *Listeria monocytogenes*. *Infect Immun* 87:e00758-18. <https://doi.org/10.1128/IAI.00758-18>.
- Molloy EM, Cotter PD, Hill C, Mitchell DA, Ross RP. 2011. Streptolysin S-like virulence factors: the continuing saga. *Nat Rev Microbiol* 9:670–681. <https://doi.org/10.1038/nrmicro2624>.
- Kuenne C, Billion A, Mraheil MA, Strittmatter A, Daniel R, Goesmann A, Barbuddhe S, Hain T, Chakraborty T. 2013. Reassessment of the *Listeria monocytogenes* pan-genome reveals dynamic integration hotspots and

- mobile genetic elements as major components of the accessory genome. *BMC Genomics* 14:47. <https://doi.org/10.1186/1471-2164-14-47>.
21. Cheng Y, Kim J-W, Lee S, Siletzky RM, Kathariou S. 2010. DNA probes for unambiguous identification of *Listeria monocytogenes* epidemic clone II strains. *Appl Environ Microbiol* 76:3061–3068. <https://doi.org/10.1128/AEM.03064-09>.
 22. Glaser P, Frangeul L, Buchrieser C, Rusniok C, Amend A, Baquero F, Berche P, Bloecker H, Brandt P, Chakraborty T, Charbit A, Chetouani F, Couvé E, de Daruvar A, Dehoux P, Domann E, Domínguez-Bernal G, Duchaud E, Durant L, Dussurget O, Entian KD, Fsihi H, García-del Portillo F, Garrido P, Gautier L, Goebel W, Gómez-López N, Hain T, Hauf J, Jackson D, Jones LM, Kaerst U, Kreft J, Kuhn M, Kunst F, Kurapkat G, Madueno E, Maitournam A, Vicente JM, Ng E, Nedjari H, Nordsiek G, Novella S, de Pablos B, Pérez-Díaz JC, Purcell R, Rimmel B, Rose M, Schlueter T, Simoes N, et al. 2001. Comparative genomics of *Listeria* species. *Science* 294:849–852. <https://doi.org/10.1126/science.1063447>.
 23. Meza-Torres J, Lelek M, Quereda JJ, Sachse M, Manina G, Ershov D, Tinevez J-Y, Radoshevic L, Maudet C, Chaze T, Giai Gianetto Q, Matondo M, Lecuit M, Martin-Verstraete I, Zimmer C, Bierre H, Dussurget O, Cossart P, Pizarro-Cerdá J. 2021. Listeriolysin S: a bacteriocin from *Listeria monocytogenes* that induces membrane permeabilization in a contact-dependent manner. *Proc Natl Acad Sci U S A* 118:e2108155118. <https://doi.org/10.1073/pnas.2108155118>.
 24. Lan Z, Fiedler F, Kathariou S. 2000. A sheep in wolf's clothing: *Listeria innocua* strains with teichoic acid-associated surface antigens and genes characteristic of *Listeria monocytogenes* serogroup 4. *J Bacteriol* 182:6161–6168. <https://doi.org/10.1128/JB.182.21.6161-6168.2000>.
 25. Lee S, Parsons C, Chen Y, Hanafy Z, Brown E, Kathariou S. 2021. Identification and Characterization of a novel genomic island harboring cadmium and arsenic resistance genes in *Listeria welshimeri*. *Biomolecules* 11:560. <https://doi.org/10.3390/biom11040560>.
 26. Parsons C, Niedermeyer J, Gould N, Brown P, Strules J, Parsons AW, Bernardo Mesa-Cruz J, Kelly MJ, Hooker MJ, Chamberlain MJ, Olfenbuttel C, DePerno C, Kathariou S. 2020. *Listeria monocytogenes* at the human-wildlife interface: black bears (*Ursus americanus*) as potential vehicles for *Listeria*. *Microb Biotechnol* 13:706–721. <https://doi.org/10.1111/1751-7915.13509>.
 27. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
 28. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. 2015. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods* 8:12–24. <https://doi.org/10.1039/C5AY02550H>.
 29. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
 30. Nelson KE, Fouts DE, Mongodin EF, Ravel J, DeBoy RT, Kolonay JF, Rasko DA, Angiuoli SV, Gill SR, Paulsen IT, Peterson J, White O, Nelson WC, Nierman W, Beanan MJ, Brinkac LM, Daugherty SC, Dodson RJ, Durkin AS, Madupu R, Haft DH, Selengut J, Van Aken S, Khouri H, Fedorova N, Forberger H, Tran B, Kathariou S, Wonderling LD, Uhlich GA, Bayles DO, Luchansky JB, Fraser CM. 2004. Whole genome comparisons of serotype 4b and 1/2a strains of the food-borne pathogen *Listeria monocytogenes* reveal new insights into the core genome components of this species. *Nucleic Acids Res* 32:2386–2395. <https://doi.org/10.1093/nar/gkh562>.
 31. Halbedel S, Prager R, Banerji S, Kleta S, Trost E, Nishanth G, Alles G, Hölzel C, Schlesiger F, Pietzka A, Schlüter D, Flieger A. 2019. A *Listeria monocytogenes* ST2 clone lacking chitinase ChiB from an outbreak of non-invasive gastroenteritis. *Emerg Microbes Infect* 8:17–28. <https://doi.org/10.1080/22221751.2018.1558960>.
 32. Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, Chlenski P, Conrad N, Dickerman A, Dietrich EM, Gabbard JL, Gerdes S, Guard A, Kenyon RW, Machi D, Mao C, Murphy-Olson D, Nguyen M, Nordberg EK, Olsen GJ, Olson RD, Overbeek JC, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomas C, VanOeffelen M, Vonstein V, Warren AS, Xia F, Xie D, Yoo H, Stevens R. 2020. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Res* 48:D606–D612. <https://doi.org/10.1093/nar/gkz943>.
 33. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Pribelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA. 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol* 20:714–737. <https://doi.org/10.1089/cmb.2013.0084>.
 34. Stothard P, Grant JR, Van Domselaar G. 2019. Visualizing and comparing circular genomes using the CGView family of tools. *Brief Bioinform* 20:1576–1582. <https://doi.org/10.1093/bib/bbx081>.
 35. Liu B, Zheng D, Jin Q, Chen L, Yang J. 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res* 47:D687–D692. <https://doi.org/10.1093/nar/gky1080>.
 36. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25:1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>.
 37. Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16:276–277. [https://doi.org/10.1016/S0168-9525\(00\)02024-2](https://doi.org/10.1016/S0168-9525(00)02024-2).
 38. da Veiga Leprevost F, Grüning BA, Alves Aflitos S, Röst HL, Uszkoreit J, Barsnes H, Vaudel M, Moreno P, Gatto L, Weber J, Bai M, Jimenez RC, Sachsenberg T, Pfeuffer J, Vera Alvarez R, Griss J, Nesvizhskii AI, Perez-Riverol Y. 2017. BioContainers: an open-source and community-driven framework for software standardization. *Bioinformatics* 33:2580–2582. <https://doi.org/10.1093/bioinformatics/btx192>.
 39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
 40. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hungwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A. 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res* 48:D265–D268. <https://doi.org/10.1093/nar/gkz991>.
 41. Solovyev V, Salamov A. 2011. Automatic annotation of microbial genomes and metagenomic sequences, p 61–78. In Li RW (ed), *Metagenomics and its applications in agriculture, biomedicine and environmental studies*. Nova Science Publishers, Hauppauge, NY.
 42. Amin MR, Yurovsky A, Chen Y, Skiena S, Futcher B. 2018. Re-annotation of 12,495 prokaryotic 16S rRNA 3' ends and analysis of Shine-Dalgarno and anti-Shine-Dalgarno sequences. *PLoS One* 13:e0202767. <https://doi.org/10.1371/journal.pone.0202767>.
 43. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539. <https://doi.org/10.1038/msb.2011.75>.
 44. Gardner SN, Slezak T, Hall BG. 2015. kSNP3: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* 31:2877–2878. <https://doi.org/10.1093/bioinformatics/btv271>.
 45. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549. <https://doi.org/10.1093/molbev/msy096>.
 46. Yu G. 2020. Using ggtree to visualize data on tree-like structures. *Curr Protoc Bioinforma* 69:e96. <https://doi.org/10.1002/cpbi.96>.
 47. Zhou Z, Alikhan N-F, Sergeant MJ, Luhmann N, Vaz C, Francisco AP, Carriço JA, Achtman M. 2018. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res* 28:1395–1404. <https://doi.org/10.1101/gr.232397.117>.
 48. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. *Bioinformatics* 27:1009–1010. <https://doi.org/10.1093/bioinformatics/btr039>.